DOCUMENT RESUME

ED 191 922 TM 800 548

AUTHOR Wainer, Howard
TITLE Some Supporting Evidence for Accurate Multivariate Perceptions with Chernoff Faces, Project 547.
INSTITUTION Bureau of Social Science Research, Inc., Washington, D.C.
SPONS AGENCY National Science Foundation, Washington, D.C.
REPORT NO BSSR-0538-33
PUB DATE Jan 79
GRANT SOC76-17768-AC1
NOTE 16p.
AVAILABLE FROM Bureau of Social Science Research, Inc., 1990 M St., N.W., Washington, DC 20036 ($1.60).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cartoons: *Factor Analysis: Higher Education: *Multivariate Analysis: *Perception Tests: *Psychometrics: Visual Perception
IDENTIFIERS *Facial Expressions: *Facial Features

ABSTRACT

A scheme, using features in a cartoon-like human face to represent variables, is tested as to its ability to graphically depict multivariate data. A factor analysis of Harman's "24 Psychological Tests" was performed and yielded four orthogonal factors. Nose width represented the loading on Factor 1: eye size on Factor 2: curve of mouth on Factor 3: and height of eyes on Factor 4. The communality (i.e. the proportions of the total variance accounted for by the factor space) was represented by the size (area) of the face. Ten "experts" who knew about Chernoff faces, factor analysis, and Harman's 24 Psychological Tests, and 32 randomly selected "novices" sorted 24 squares depicting "faces" into as many piles as they thought sensible. Each subject's responses were coded, and group similarity matrices were constructed for the experts and the novices by averaging across their individual matrices. Results indicated that, for a four dimensional structure, untrained human judges performed with astonishing accuracy. "Faces" not only allowed the clustering of multivariate data points perceptually, but the distances between points as well. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191922

Project 547
BSSR 0538-33

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SOME SUPPORTING EVIDENCE FOR ACCURATE MULTIVARIATE PERCEPTIONS WITH CHERNOFF FACES

by

Howard Wainer

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
P. TAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This research was funded by a grant from the National Science Foundation (SOC76-17768 A01) to the Bureau of Social Science Research. Albert D. Biderman and Howard Wainer, project directors. The author thanks Albert D. Biderman, Curtis Groves, Marianne Lono, Roderick McDonald and Douglas Neal for their help and suggestions. This paper is an expansion of a talk given at the Psychometric Society meetings in Hamilton, Ontario, in August 1978.

Bureau of Social Science Research, Inc.
1990 M Street, N.W.
Washington, D.C. 20036

January 1979

TM 800548

Some Supporting Evidence for Accurate Multivariate Perceptions with Chernoff Faces

ABSTRACT

Chernoff (1) has proposed the use of cartoon-like human faces to aid in the graphical depiction of multivariate data. An experiment was undertaken which tested this proposal, and which found that for a four dimensional structure untrained human judges performed with astonishing accuracy.It is concluded that Faces not only allow the clustering of multivariate data point perceptually, but that the distances between points is accurately perceived as well.

I. Introduction

A variety of schemes to aid in the graphical depiction of multivariate data have been proposed recently. Bachi (2) uses patterns of dots which grow to squares as one method. Andrews (3) has proposed representing each data point as a periodic function, in which each coordinate of that point in a multidimensional space is a Fourier component. Chambers (4) supports star-like structures, in which the length of each arm of the star is proportional to a corresponding component of the data point. One of the most charming, and indeed, according to one investigator (Jacob, 5), the most efficacious, is the Chernoff (1) "FACE".

Chernoff felt (and there is ample experimental evidence to support him) that humans are very good indeed at seeing rather subtle differences in human faces, at grouping faces by certain global characteristics which are a gestalt of many features, and at remembering the many features of the face. This latter ability may exist partially because we have names for particular aggregations of characteristics, i.e. happy oriental, glum Cyrano, innocent nit-wit, etc. The logic then proceeds: why not represent each variable as a feature in a cartoon face? The whole face will then be the multivariate representation of that data point. This sort of representation, if done wisely, will allow each variable to be separated out from the whole, but the variable will also be integrated into a holistic perception and be amenable to perceptual clustering. Thus Variable One might be "eye-size" with a high score having big eyes and a low score having small eyes. Variable Two might be "mouth curve" with a high score being a big smile and a low score a

big frown. The third variable might be nose size, and so on. It only remains to use our imaginations to devise a large number of features for variable representation and to test how well this featural representation works.

The efficacious use of Chernoff faces would be greatly aided if we knew how well the multivariate structure being portrayed by the faces was conveyed with integrity to human observers. An experiment was performed to try to answer this question.

II. The Experiment

A data set which has become a classic example in the field of factor analysis was used to generate the stimuli. This data set is Harman's (6, p. 164) "24 Psychological Tests". Rather than depict the raw scores of each test for each subject, which is a problem of much larger extent than is practical to deal with experimentally, we used the results of a factor analysis of these data. This analysis, which is used as an example in most courses on factor analysis, yields four orthogonal factors. Chernoff (personal communication) has suggested that faces work best when the variables being represented are orthogonal, and four dimensions are quite large enough to present a challenging depiction problem for most standard graphical techniques.

The coordinates of the 24 tests in four-space are shown in Table 1, along with their associated communalities. Figure 1 presents the faces drawn from these data points. The nose width represents the loading on Factor 1, the eye size represents the loading on Factor 2, the curve of the mouth represents the loading on Factor 3, and the height of the eyes represents the loading on Factor 4. In addition, we represented the communality (the proportion of the total variance accounted for by the factor space) by the size (area) of the face. It was felt that this was an appropriate visual metaphor for this variable.

Insert Table 1 and Figure 1 About Here

The Subjects

There were two groups of subjects. First, there were ten

"experts" who knew about Chernoff faces, were familiar with factor analysis in general, and this data set in particular. They are all leading figures in the field of psychometrics. The second group was made up of 32 "novices", including randomly selected undergraduates at Temple University, non-professional staff at BSSR, and spouses of other participants. They knew nothing of factor analysis, Chernoff faces, nor Harman's 24 Psychological Tests.

The Experimental Procedure

The 24 faces as shown in Figure 1 were cut-up into 24 squares (3 cm. on a side) and placed into envelopes. The subjects were instructed to sort the faces into as many piles as they thought sensible, grouping together those they felt belonged in the same pile. The task was not speeded.

The Analysis and Results

Each subject's responses were coded into a 24-by-24 matrix, X, with entries x_{ij}, where x_{ij} was 1 if face i was grouped with face j, and zero otherwise. A group similarity matrix was formed by averaging across these individual matrices, yielding a matrix whose entries ranged from zero (when two faces were never grouped together by anyone), to one (when two faces were always grouped together). One matrix was formed for each of the two subject groups.

Each similarity matrix was analyzed using a standard multidimensional scaling algorithm (7) and both yielded a four dimensional solution. It is difficult to compare these two experimentally derived configurations in four space with the original data matrix (Table 1), since the solutions are specified only up to a rotation and scale factor. We therefore rotated each of the three

configurations to fit as closely as possible (in a least squares sense) what is accepted as the "Classic Solution" (6, p. 289). To facilitate comparison further we scaled the length of each data point vector to unity. The three resulting 24-by-4 matrices of item coordinates were truncated to one decimal place (and the decimal omitted) and are shown in Table 2.

Insert Table 2 and Figure 2 About Here

The pattern of loadings which are typically thought to be the "important" ones have been boxed and separated for ease of viewing. As is evident, the similarity of the three solutions (two of which were 'filtered' through Chernoff faces and human perceptions) is astonishing. The ease with which these results can be viewed will be aided if we represent these three tables with Faces. These are shown in Figure 2. The faces are constructed using the same rules as for Figure 1, except that since we have standardized the length of each point vector, all faces are the same size. Now the similarity of the three solutions is visually evident, as are those situations in which there was some deviation (note Face 10, in which the eyes {Factor 2} seem to be underestimated by the subjects).

III. Discussion

This study has shown that Chernoff faces can represent a multivariate structure of high dimensionality in a way which not only allows the accurate clustering of points in that space, but also in a way which yields an accurate perception of the relative distances between the points in that space. It appears that training does not seem to have any effect, in that a group of novice subjects did about as well as a group of experts. This implies that for many practical problems of the depiction of multivariate data points, Chernoff faces may provide a solution. For example, in educational assessment the current mood is in favor of multiple assessments being reported rather than collapsing into a single score. To do this effectively a means of communicating this multivariate assessment to untrained individuals (i.e. parents of school children) must be found. Faces seem a likely candidate.

Of course there are still many questions that are unresolved. First, our experiment averaged across individuals rather than obtaining similarity measures from just one. Does this accurate perception of the multivariate structure observed in the group exist for an individual? The accurate grouping we obtained with five variables generalize to six or seven, or how far? How much deterioration in accuracy is seen if the variables representing the data points are not orthogonal? How much of a change in perceptual accuracy is seen if we use a different featural presentation? These await further work, but so far faces are doing very well indeed.

IV. References

1. Chernoff, H. (1973) The use of faces to represent points in k-dimensional space graphically. Journal of the American Statistical Association, 68, 361-368.

2. Bachi, R. (1968) Graphical Rational Patterns: A New Approach to Graphical Presentation of Statistics. Jerusalem: Israel Universities Press.

3. Andrews, D. F. (1972) Plots of high-dimensional data. Biometrics, 28, 125-136.

4. Chambers, J. (1978) "Methods for displaying multivariate data". A talk given at the First Conference on Social Graphics. Leesburg, Virginia.

5. Jacob, R. J. K. (1976) Computer-Produced Faces as an Iconic Display for Complex Data. Unpublished Ph. D. Dissertation, The Johns Hopkins University.

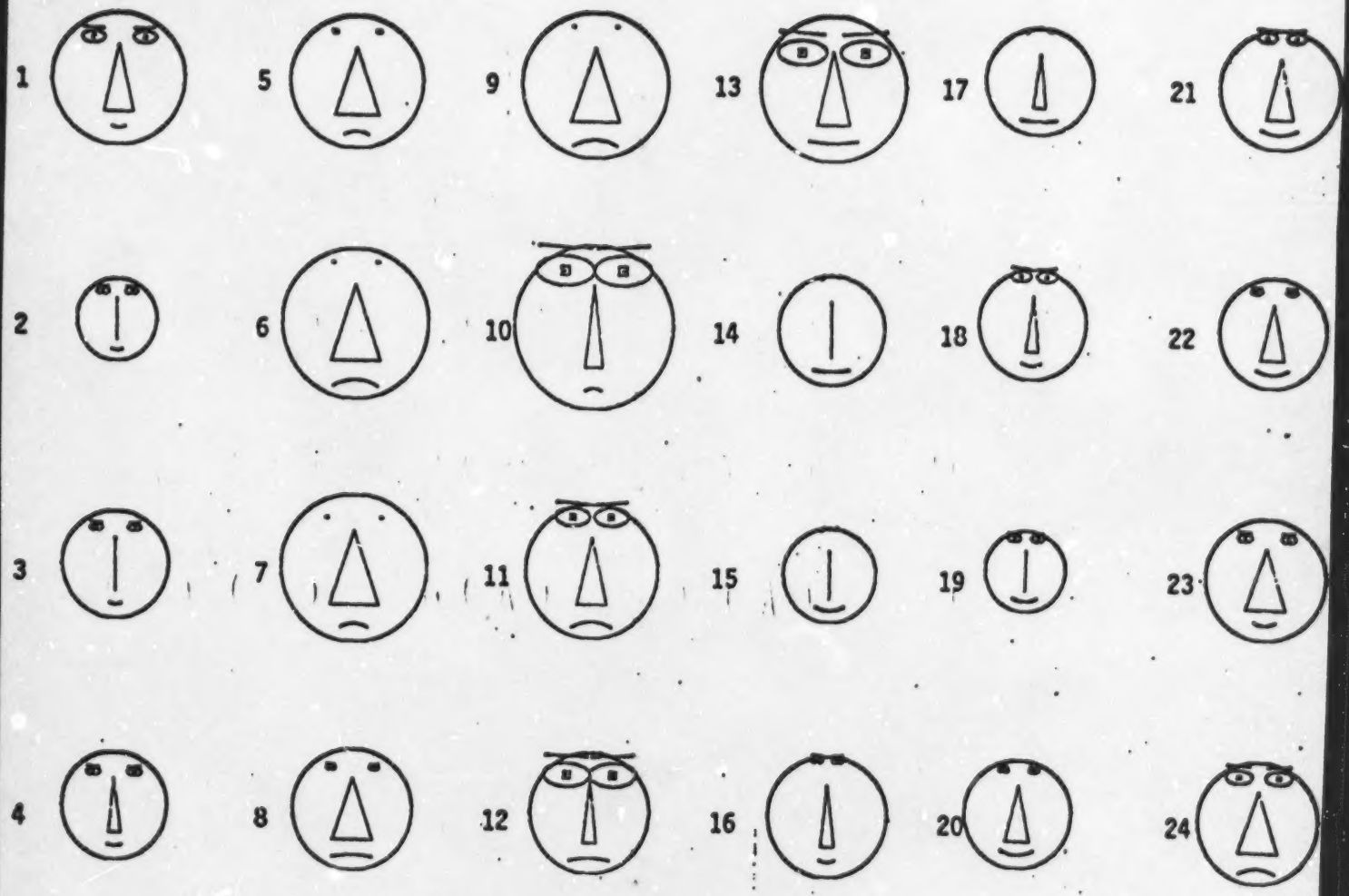
6. Harman, H. H. (1976) Modern Factor Analysis (third edition). Chicago: University of Chicago Press.

7. Kruskal, J. B. (1964) Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. Psychometrika, 29, 1-28.

Figure Captions

Figure 1. Twenty-four Psychological Tests - Unrotated Solution (Harman, 1976, p. 164) Factor 1 = Nose Width Factor 2 = Eye Size Factor 3 = Curvature of Mouth Factor 4 = Eye Height Communality = Head Size

Figure 2. Twenty-Four Psychological tests with three solutions. Rotated from Harman (1976, p.289) The other two solutions from the experiment with expert and novice subjects. Factor 1 = Nose Width Factor 2 = Eye Size Factor 3 = Mouth Curve Factor 4 = Eye height


1
5
9
13
17
21
2
6
10
14
18
22
3
7
11
15
19
23
4
8
12
16
20
24

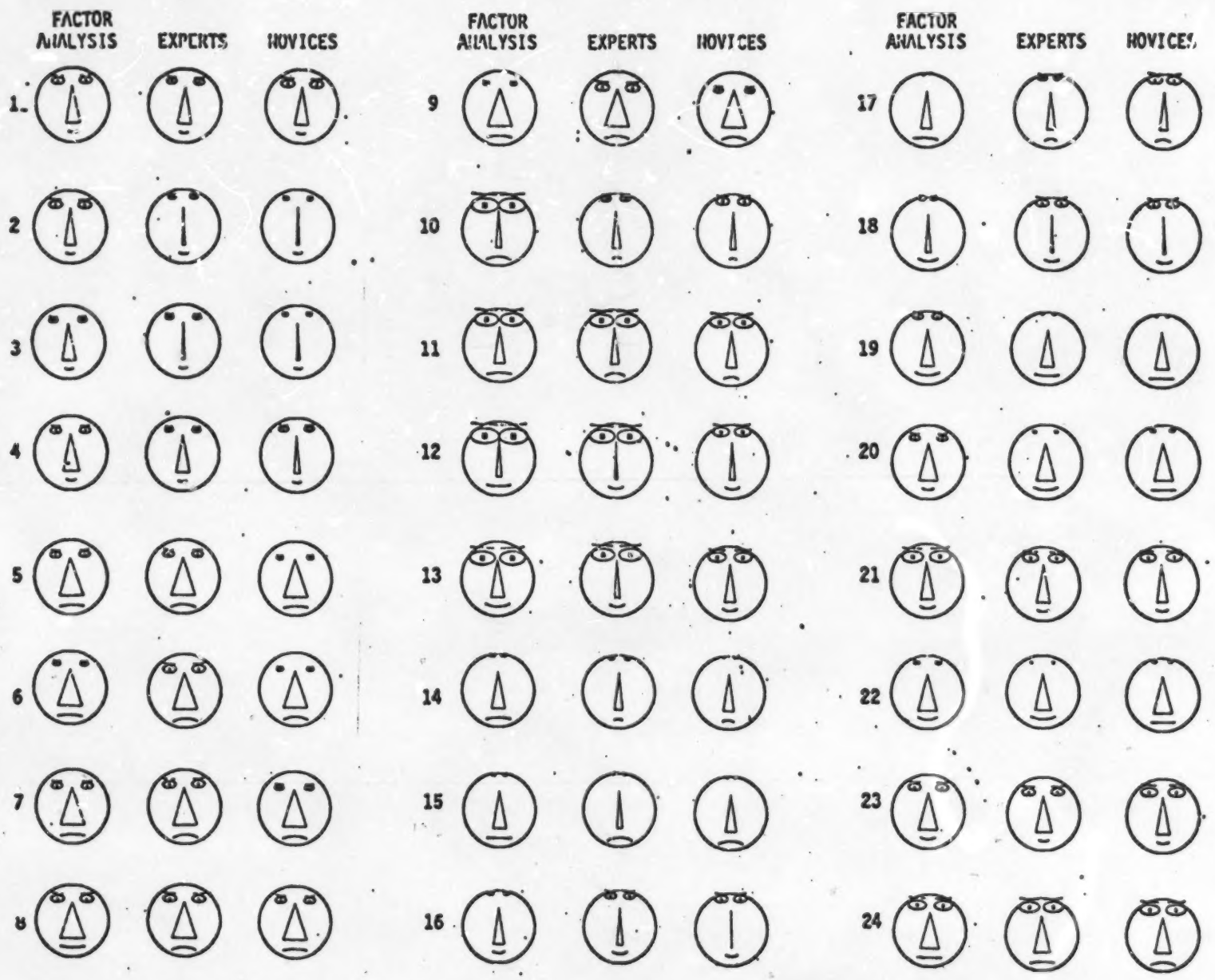

FACTOR ANALYSIS
EXPERTS
NOVICES
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Table 1. The Coordinates of 24 Psychological Tests in Four Dimensions (adapted from Harman, 1976)

Test	Common Factors				Communality
	P_1	P_2	P_3	P_4	
1	.595	−.039	.369	−.184	.531
2	.374	.026	.270	−.147	.250
3	.433	.115	.396	−.112	.446
4	.501	.108	.290	−.178	.380
5	.701	.312	−.273	−.050	.666
6	.683	.404	−.213	.067	.690
7	.676	.412	−.284	−.082	.716
8	.680	.206	−.088	−.115	.540
9	.690	.446	−.212	.076	.727
10	.456	−.469	−.446	−.128	.654
11	.589	−.372	−.198	.076	.565
12	.448	−.491	−.154	−.263	.537
13	.590	−.268	.018	−.300	.575
14	.435	−.063	−.012	.418	.371
15	.390	−.102	.055	.362	.307
16	.512	−.098	.325	.259	.444
17	.471	−.212	−.036	.388	.426
18	.521	−.331	.118	.145	.417
19	.450	−.115	.110	.167	.287
20	.623	.135	.142	.049	.492
21	.596	−.220	.076	−.140	.471
22	.600	.103	.138	.053	.413
23	.685	.063	.160	−.096	.532
24	.635	−.169	−.192	−.009	.475
V_j	7.665	1.672	1.208	.920	11.912
$100V_j/11.943$	64.2	14.0	10.1	7.7	99.7

Source: Harman, H. (1976) Modern Factor Analysis, Third Edition Chicago, University of Chicago Press. P. 194.

Table 2. A Comparison of the Rotated Item Coordinates For Three Methods
(All item vectors are normalized to unit length)

Item	Factor Loadings (Harmon, P. 289)				Expert Judges (N=10)				Novice Judges (N=32)			
1	5	2	**7**	1	5	1	**8**	1	5	2	**8**	0
2	4	2	**8**	0	1	1	**8**	4	1	-1	**8**	3
3	5	0	**8**	0	1	0	**9**	1	1	-1	**8**	3
4	6	1	**7**	0	5	0	**8**	1	2	1	**8**	2
5	**9**	1	0	0	**9**	1	-1	0	**9**	-1	0	-1
6	**9**	0	0	1	**9**	2	-1	0	**9**	-1	-1	-1
7	**9**	1	0	-1	**9**	2	-1	0	**9**	0	-1	-1
8	**9**	2	2	0	**9**	2	-1	0	**9**	1	0	0
9	**9**	-1	0	1	**9**	2	-1	-1	**9**	0	-1	-2
10	2	**9**	-1	1	3	**3**	-6	5	2	**4**	-7	4
11	4	**8**	0	2	3	**8**	-2	2	4	**7**	-3	3
12	2	**9**	2	0	1	**8**	5	1	2	**6**	5	4
13	5	**7**	3	-1	2	**7**	5	1	4	**5**	5	2
14	5	3	0	**7**	3	1	-6	**6**	4	-2	-4	**7**
15	5	1	1	**7**	2	-4	-1	**8**	4	-2	-1	**8**
16	4	1	6	**6**	2	4	5	**5**	0	5	6	**5**
17	4	2	0	**8**	3	4	-4	**6**	3	6	-3	**5**
18	2	4	4	**7**	1	5	5	**5**	1	5	5	**5**
19	5	3	3	**5**	5	0	2	**7**	5	0	1	**7**
20	**7**	1	6	1	**7**	-2	3	4	**7**	0	1	5
21	**5**	6	5	1	**5**	4	6	2	**4**	4	6	2
22	**7**	0	4	4	**6**	-3	3	4	**7**	0	1	6
23	**7**	2	5	1	**6**	2	7	0	**5**	4	6	1
24	**7**	5	1	2	**6**	7	0	3	**6**	6	-1	2